

Journal of Family Strengths

Volume 17

Issue 1 *Innovative Practices to Eliminate Health Disparities*

Article 4

5-10-2017

A Novel Bioinformatic Approach to Understanding Addiction

Latifa F. Jackson

Howard University, latifa.f.jackson@gmail.com

Follow this and additional works at: <http://digitalcommons.library.tmc.edu/jfs>

Recommended Citation

Jackson, Latifa F. (2017) "A Novel Bioinformatic Approach to Understanding Addiction," *Journal of Family Strengths*: Vol. 17 : Iss. 1 , Article 4.

Available at: <http://digitalcommons.library.tmc.edu/jfs/vol17/iss1/4>

The *Journal of Family Strengths* is brought to you for free and open access by CHILDREN AT RISK at DigitalCommons@The Texas Medical Center. It has a "cc by-nc-nd" Creative Commons license (Attribution Non-Commercial No Derivatives) For more information, please contact digitalcommons@exch.library.tmc.edu



A Novel Bioinformatic Approach to Understanding Addiction

Acknowledgements

LFJ would like to acknowledge Dr. Ceylan Tanes, Maksim Shestov and Dr. Aydin Tozeren for important conversation on the bioinformatics implemented in this project.

Introduction

Substance abuse disorders incur a significant cost to global judicial and healthcare systems (Lynskey & Strang, 2013). Addiction is loosely defined as a chronic relapsing spectrum disorder characterized by a loss of control over substance use (Girard & Carlton, 1978; Goodman, 1990; Peele, 1977). It is a behavior-based phenomenon representing a diverse array of psychological, biological, and genetic attributes, along with environmental and cultural factors (Goodwin, 1975; Naranjo & Bremner, 1993; Truan, 1993). Additionally, individuals addicted to illicit substances are stigmatized, with a widespread marginalization of rehabilitation and recovery services that perhaps facilitates recidivism (Cunningham, Sobell, & Chow, 1993; Dean & Rud, 1984; Room, 2005). The current standard of care for individuals attempting to overcome addiction behaviors emphasizes complete abstinence from addictive substances, and to a lesser extent treatment (Hoffman & Goldfrank, 1990; Lorman, 2013).

A large number of studies, including genome-wide association investigations, have uncovered potentially relevant allelic contributors to the genetic and molecular basis of addiction phenotypes (Kreek, Nielsen, & LaForge, 2004; Le Foll, Gallo, Le Strat, Lu, & Gorwood, 2009; Messer et al., 2000; Olfson & Bierut, 2012; Pependikyte et al., 1999; Shi et al., 2002; Treutlein & Rietschel, 2011). Genetic studies such as those conducted on the alcohol dehydrogenase gene (*ADH*) family have been important for advancing our understanding of the genetic basis of addiction phenotypes (Crabb, Bosron, & Li, 1987; Pietruszko, 1975). The literature describes two main molecular areas underlying addiction phenotypes. The first is analyzed at the level of neurological function (Chao & Nestler, 2004; Gardner, 2011; Maldonado, 2003). Disruption of the normal range of neurological function has been well characterized both in the literature and in the canonical metabolic pathways that have been identified as participating in addictions such as alcohol, cocaine, and dopaminergic substance abuse (Commons, 2010; Hillemecher, 2011; Sun & Zhao, 2010). The second is addressed at the level of metabolic function (Chiang, Lo, & Chen, 2013), such as characterization of the *ADH* gene family and its function in the metabolism of alcohol products (Cotton & Goldman, 1988; Ehlers, Liang, & Gizer, 2012).

Association studies are yet to provide a complete picture for elucidating the complex pathways and mechanisms underlying addiction phenotypes. The population subtype-dependent mechanisms of addiction have not been fully explored (Pasche & Yi, 2010; Uhl et al., 2008; Yuferov, Levran, Proudnikov, Nielsen, & Kreek, 2010), and neither have the epigenetic factors involved in addiction (Petronis, 2010; Ponomarev, 2013;

Schumacher & Petronis, 2006; R. Zhang et al., 2013). Because the tissues involved in addiction, particularly neurons and liver tissue, cannot be experimentally investigated in humans, experiments in animal models and in vitro experiments have been used to quantify epigenetic, transcriptomic, metabolomic, and proteomic changes in addiction-perturbed tissues (Adkins et al., 2013; Brown et al., 2012; Contet, 2012; Freeman et al., 2005; Lee & Messing, 2008; K. W. Li et al., 2006; Mulligan et al., 2006; Saito et al., 2004; G. C. Zhang et al., 2009). This has limited our ability to understand how addiction gene variation plays out in the geographically and ethnically diverse human populations that are afflicted by addiction disorders. Studying a diverse array of ethnic populations that represent major human geographical regions and ancestries would be particularly useful in understanding the role that variation plays at potential addiction complex sites in the genome.

Opiate, dopamine, and gamma aminobutyric acid (GABA) addictions are complex diseases with strong genetic components. Addiction to opioids, amphetamines, and alcohol arise owing to a complex array of social, genetic, and environmental factors. This paper demonstrates a comprehensive computational approach to incorporating biological data sets to address addiction in ethnic minority populations. Combining these biological data types both identifies novel, functionally relevant genetic variants for addiction and informs the conditions under which these variants are correlated to addiction phenotypes.

Methods

Addiction-Linked Genes and Genome Hotspots

The National Center for Biotechnology Information (NCBI) Gene database was queried according to the word combinations shown in Figure 1 to generate a list of genes with biological relevance to dopamine (cocaine and crystal methamphetamine), opiate (heroin and morphine), and GABA (alcohol and gamma hydroxybutyrate [GHB]) addiction phenotypes. This unique set of genes was then mapped onto human chromosomes, and clusters of genes were considered (Blankenberg et al., 2010; Giardine et al., 2005; Goecks, Nekrutenko, Taylor, & Galaxy, 2010). Browser Extensible Data (BED) output files were used to visualize the hotspot genes on the University of California, Santa Cruz (UCSC) Genome Browser as a custom track within the HG-19 human genome build.

Hotspots were defined as genic regions of approximately 1.0 to 1.5 Mb, corresponding to a human centiMorgan unit of recombination, in physical distance along the genome that contained six or more genes identified from our combined addiction gene list. Each hotspot contained

genes not currently associated with addiction phenotypes. We included such genes into our analysis because of the high probability of common regulation patterns within a hotspot. Next, we investigated the statistical significance of observing six or more genes as hotspots by using the hypergeometric test statistic for chromosomal hotspot gene enrichment.

Functional Annotation of Hotspot Genes

Genes located within hotspots were considered in two ways in statistical enrichments: all genes in the hotspot window and only those previously linked to addiction. All genes in the hotspots were annotated with tools in the DAVID (Database for Annotation, Visualization, and Integrated Discovery) Bioinformatics Resources software (Huang da, Sherman, & Lempicki, 2009; Huang da et al., 2009) for biological process, molecular function, and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (Kanehisa, 2002; Kanehisa, Goto, Kawashima, & Nakaya, 2002). Functional enrichments were quantified with Benjamini score analysis cutoffs of 0.01 (Benjamini, Drai, Elmer, Kafkafi, & Golani, 2001). A MATLAB code was written to multicolor the nodes in KEGG pathways to differentiate between genes belonging to different hotspots.

Genetic Variation in Hotspots in Population Subtypes

We examined hotspot-associated polymorphisms identified in 11 HapMap sample populations with distinct geographical distributions. These were East Asian ancestry [Japanese-JPT, Chinese (collected in Beijing)-CHB, Chinese (collected in Denver)-CHD]; African ancestry [Yorubans-YRI, Masaai-MKK, Luhya-LWK, and African Americans-ASW]; European ancestry [Europeans of Northern and Western Ancestry-CEU and Toscana-TSI]; South Asian ancestry [Gujaratis in Houston-GIH]; and an admixed American population [Mexicans in Los Angeles-MEX] (International HapMap Consortium, 2005; International HapMap Consortium et al., 2010; International HapMap Consortium et al., 2007). To exclude the possibility of confounding effects of population-specific demography and to set up an empirically derived neutral estimate of allelic variation, we analyzed 20 concatenated autosomal loci across the human genome identified as neutrally evolving (Wall et al., 2008). It was assumed that polymorphism variation undergoing selection will have patterns divergent from neutral allele frequency patterns. Single-nucleotide polymorphism (SNP) frequencies were trimmed to exclude SNPs that were almost fixed in populations (>0.9) or of low frequency (<0.15). Average SNP frequencies were calculated across the window and compared in all populations. The Kolmogorov-Smirnov test (Fuerst, Chakraborty, & Nei, 1977) was

performed to test for pairwise differences in the distributions of polymorphism profiles between populations at the concatenated neutral region and all addiction hotspots. Populations were clustered by regional ethnic origin: Africans [Luhya (LWK), Masaai (MKK), Yorubans (YRI), and African Americans (ASW)]; Asians [Chinese-Beijing (CHB), Chinese-Denver (CHD), and Japanese (JPT)]; and Europeans [Western Europeans (CEU) and Toscana (TSI)].

Populations with similar geographical ethnic origins were concatenated by continent of origin and then compared [African (ASW, YRI, MKK, LWK) with European (CEU, TSI); African with Asian (CHB, CHD, JPT); and European with Asian] in order to conduct pairwise assessments of significantly different polymorphisms; a chi-square test was used to test the divergence of the frequency ratios from 1. Significant polymorphisms were then annotated by their genomic location as coding, intronic, or intergenic polymorphisms. These significantly different polymorphisms were compared with known genotype-phenotype associations by using GWAS3D, a web-based software program that identifies Genome-Wide Association Study (GWAS) Central regulatory elements for long-range linkage and cross-chromosome interactions (M. J. Li, Wang, Xia, Sham, & Wang, 2013).

Results

Molecular Functions and Biological Processes of Addiction Genes

A list of biologically relevant genes was generated from searches with NCBI Gene with phenotypic relationships to opiate, dopamine, and GABA addiction. Three classes of addiction genes were thus obtained: dopamine addiction genes (N=108), opiate addiction genes, (N=246) and GABA addiction genes (N=433). Because addiction can arise from substance abuse or misregulation of metabolism, search terms were categorized as either “metabolism” (N=398 genes) or “addiction” (N=461 genes) followed by the names of three addictive substances: dopamine, opiate, and GABA receptor. The respective heat maps in Figure 1 illustrate the intersections within the search terms for both the metabolism and the addiction gene lists. Additionally, we compared our gene list with the one reported by C. Y. Li, Mao, & Wei (2008), which contains 387 genes involved in four addiction disorders. This comparison is shown in a Venn diagram, with the majority of genes (N=311) not identified in our analysis belonging to nicotine addiction, a disorder not addressed here. A set of unique addiction genes (N=587), compiled from the union of all search terms, was thus determined.

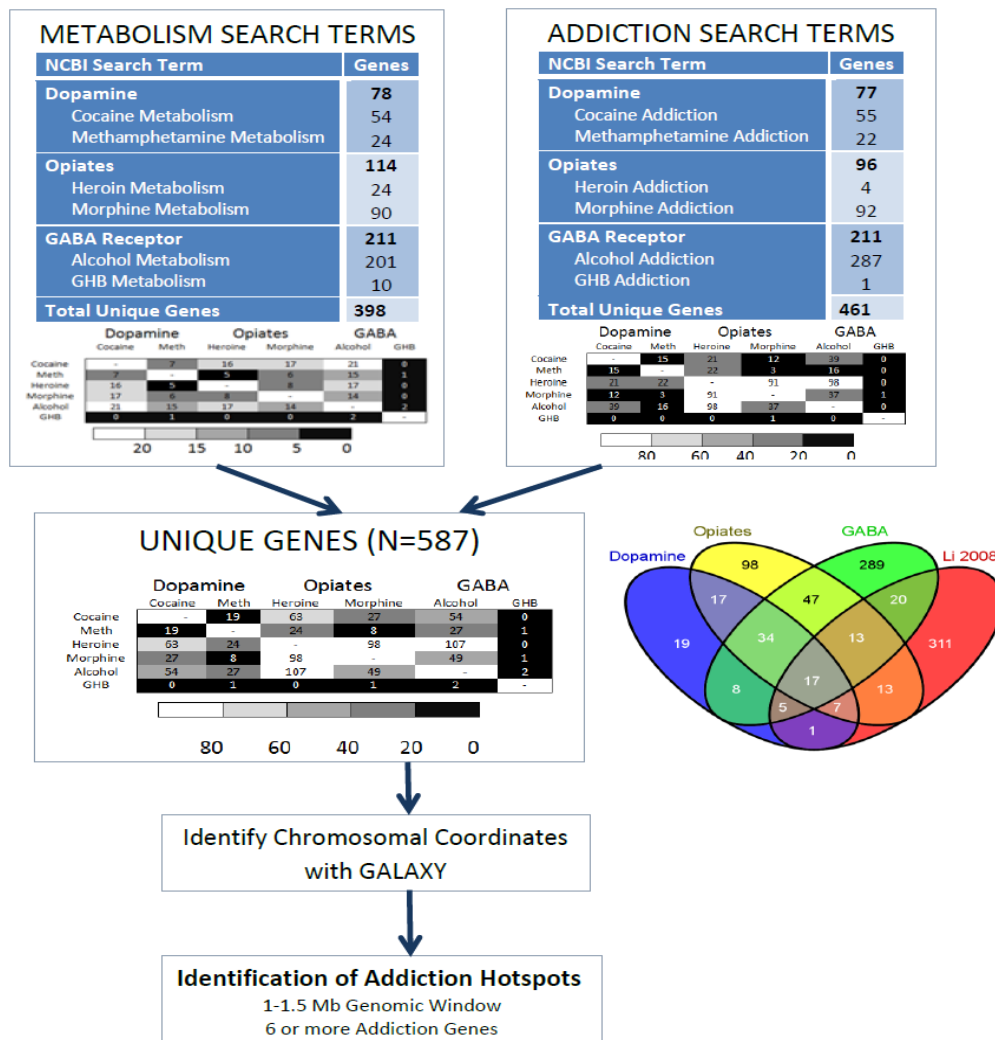


Figure 1. Flowchart for identifying biologically relevant addiction genes. The search terms used at NCBI Gene to populate a list of unique addiction genes are shown. The heat maps indicate the intersection of gene sets in three classes of addiction: dopamine, opiate, and GABA. Addiction hotspots were defined as a genomic region with six or more addiction genes within a 1- to 1.5-Mb genomic window. A Venn diagram shows the comparison with the 387 genes identified through an alternate addiction study (C. Y. Li et al., 2008).

Figure 2 illustrates the overlap in the opiate, dopamine, and GABA gene sets. Roughly 10% of addiction genes were shared among all addiction types (N=51 genes). This subset included leucine zipper family of DNA binding proteins, genes that code for glutamate-gated ion channels, and sodium:neurotransmitter symporters. Annotation of the overlapping

sets of genes for dopamine, opiates, and GABA indicate that these literature-curated addiction genes are involved in diverse processes underlying addictive xenobiotic metabolism.

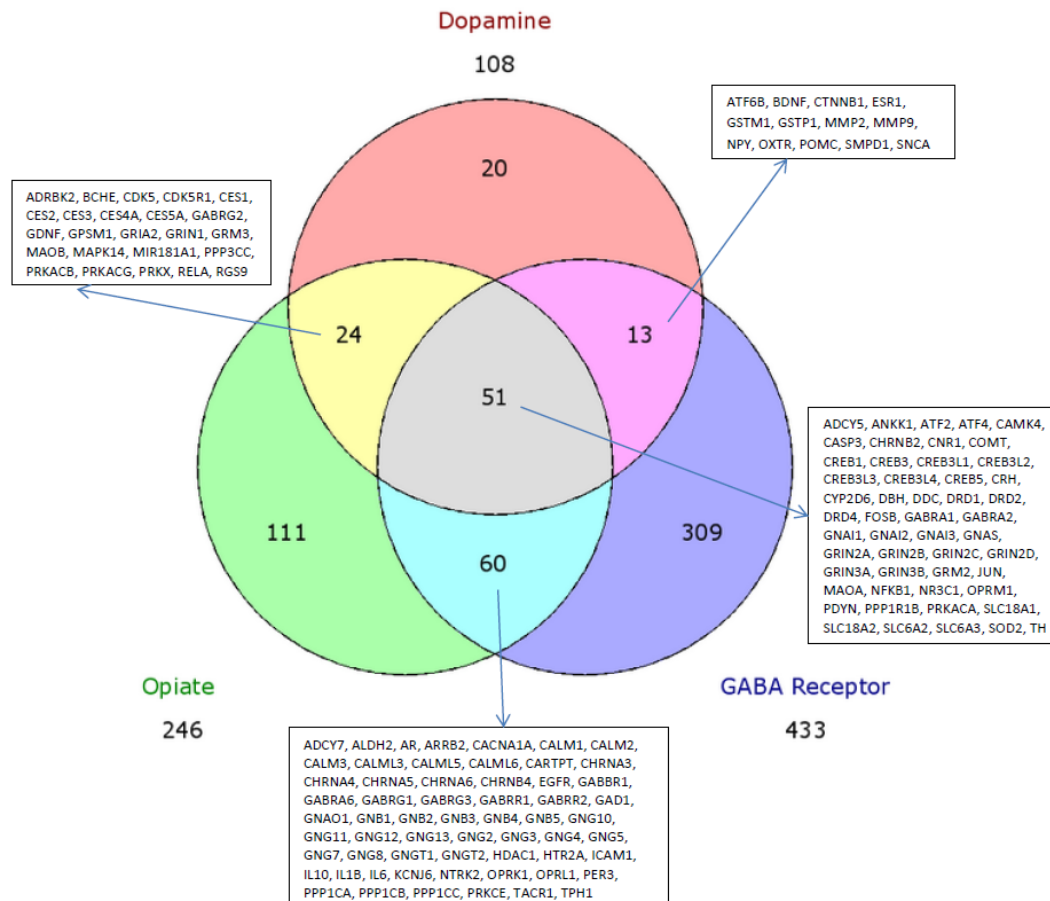


Figure 2. Genes shared by dopamine, opiate, and GABA addiction sets. The figure provides the gene symbols for the intersections of gene sets corresponding to dopamine, opiate, and GABA addictions.

Identification and Annotation of Addiction Hotspots on the Human Genome

Unique addiction genes were mapped onto the human genome; 63% of the unique set of addiction genes were mapped onto seven genomic hotspot regions, all less than 1.5 Mb, the typical length for genetic variation in human chromosomes (Figure 3). All these hotspots had at least six or more addiction genes. A hypergeometric test was performed to determine the statistical significance of all genomic hotspots.

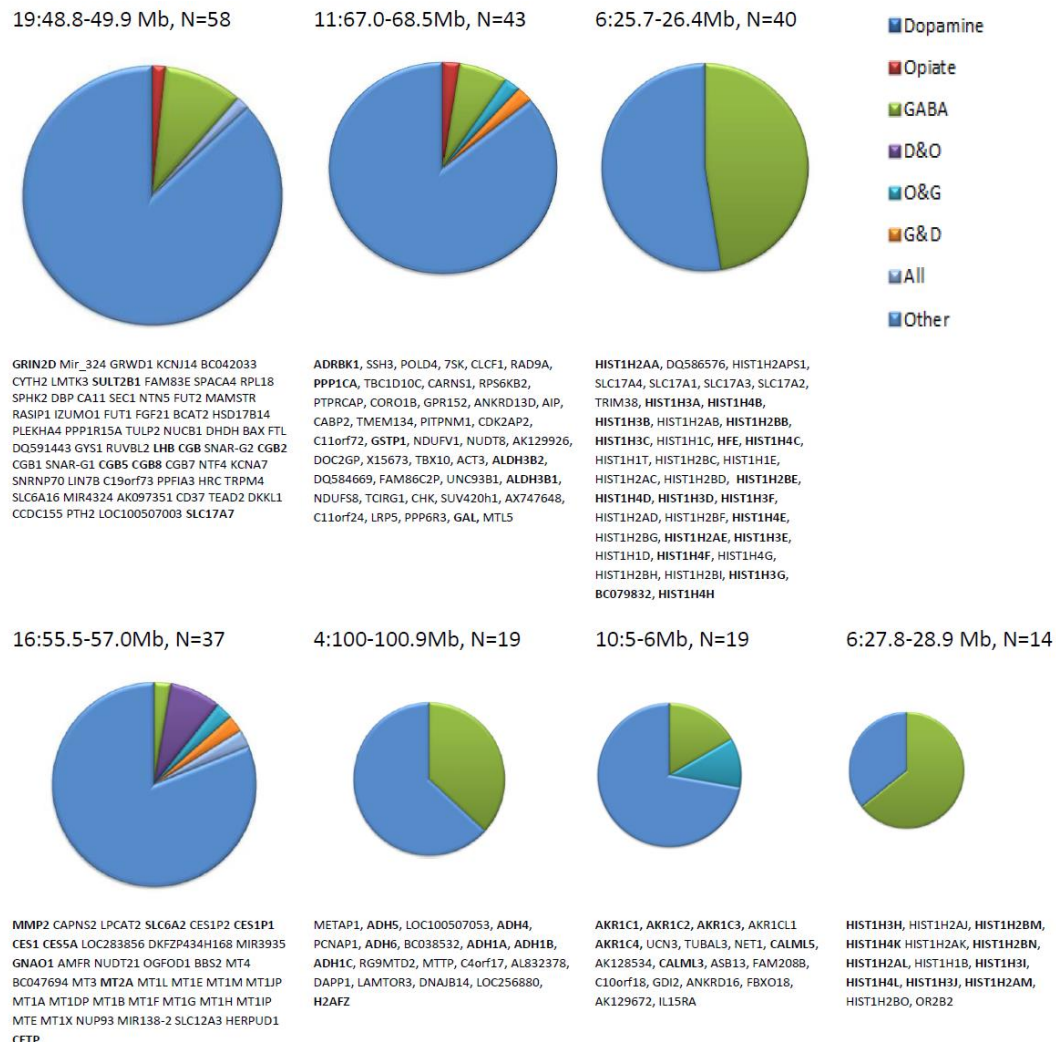


Figure 3. Seven addiction hotspots identified on the human genome. Each pie chart represents a hotspot, with the size proportional to the total number of genes within the 1- to 1.5-Mb window. Three of the hotspots were functionally related exclusively to GABA function, whereas four hotspots were mixed in addiction activity. Genes that lie within each hotspot were identified through the UCSC Genome Browser.

Three of the seven hotspots contained genes exclusively associated GABA addiction, whereas the remaining four hotspots contained genes involved in GABA, dopamine, and opiate addictions (Figure 3). For each hotspot, we cataloged the co-located genes in the hotspots not yet identified as addiction-related. These additional genes are candidates for further investigation. Hotspot windows contained 14 to 58 genes, with the number of addiction genes ranging from 6 to 19 curated genes located in a hotspot

interval. The genes in each hotspot are shown in Figure 4, with those already linked to addiction shown in boldface type.

Region		Biological Processes		Molecular Function	
	Chr Location	Functional Annotation	Benjamini	Functional Annotation	Benjamini
4q23	4:100Mb-100.9Mb	Ethanol metabolic process	5.0E-9	Alcohol dehydrogenase activity	2.9E-7
GABA Only		Ethanol oxidation Alcohol metabolic process	5.0E-9 5.0E-9		
6p22.2	6:25.7Mb-26.4Mb	Nucleosome assembly	3.2E-9	DNA binding	5.6E-3
GABA Only		DNA packaging Chromosome organization	3.4E-9 1.8E-6	Ion membrane transporter activity Alkali metal ion binding	7.5E-5 7.5E-3
6p22.1	6:27.8Mb-28.9Mb	Nucleosome assembly	2.9E-8	DNA binding	1.0E-3
GABA Only		DNA packaging Chromosome organization	2.2E-8 1.3E-5		
10p15.1	10:5Mb-6Mb	Oxidation reduction	1.8E-2	Steroid dehydrogenase activity	1.5E-6
Mixed		Steroid metabolic process	2.8E-2	Trans- 1,2-dihydrobenzene -1,2-diol dehydrogenase activity Aldo-keto reductase activity	5.7E-6 7.5E-5
11q13.2-3	11:67Mb-68.5Mb	None	NA	Aldehyde dehydrogenase activity	3.6E-2
Mixed					
16q22.1	16:55.5Mb-57Mb	None	NA	Cadmium ion binding Cation binding Transition metal ion binding	1.8E-16 2.2E-5 2.7E-3
19q13.33	19:48.8Mb-49.9Mb	Cell-cell signaling	0.02	Hormone activity	0.03
Mixed					

Table 1. Addiction genes within hotspots show distinct functional classification. The table shows the gene ontology molecular functions and biological processes statistically enriched in addiction hotspots (Benjamini coefficient <0.03). Annotations in boldface correspond to the addiction genes identified within a cluster, not the entire genome region.

The DAVID Gene Ontology data sets were used identify biological processes and molecular functions enriched for each hotspot. Table 1 summarizes the results. The chromosome 4 hotspot for GABA addiction is

dominated by metabolic processes and enzymes. A GABA-only hotspot on chromosome 6 is crowded by genes involved in nucleosome assembly and DNA packaging. The adjacent GABA hotspot is also dominated by DNA binding proteins. The genes crowding the mixed hotspot on chromosome 10 are involved in oxidation reduction and steroid metabolic processes. This hotspot has a series of neurological function genes that regulate appetite during stress in the brain and neuro-epithelial remodelers. The chromosome 11 hotspot contains multiple genes involved in actin creation/dynamics, cell development/differentiation, and lipid metabolism, in addition to metallothionein genes involved in the metabolism of xenobiotics. The chromosome 19 hotspot contains genes mediating spermatogenesis, hormone activity, and signaling. It is clear from these results that addiction hotspots contain genes with complementary functions.

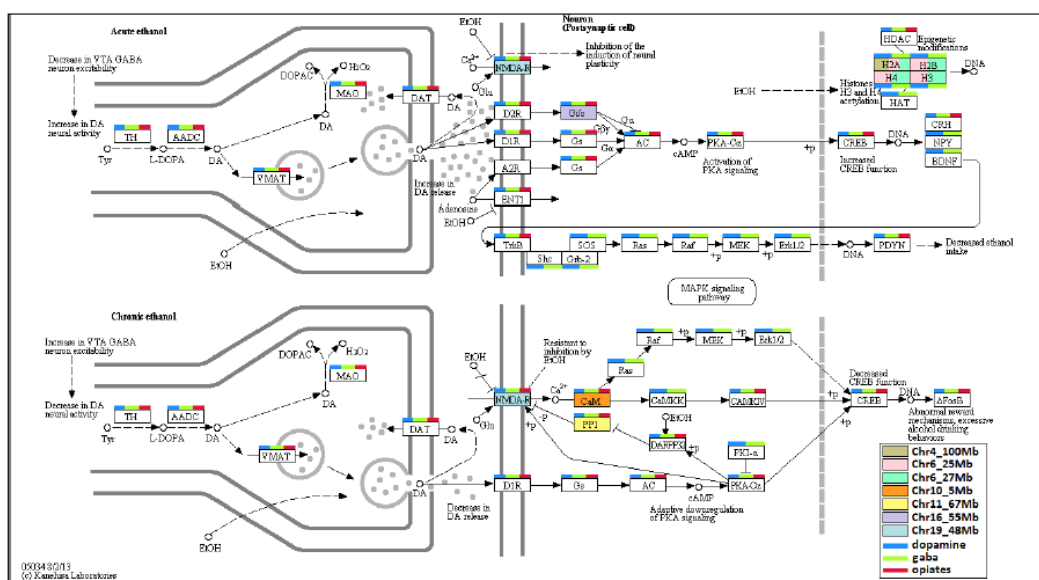


Figure 4. Hotspots participating in acute and chronic alcoholism pathways. The alcohol addiction pathway contains genes located in all seven of the addiction hotspots; each is colored according to the legend. They participate in post-synaptic processes in both acute and chronic alcohol signaling. Additionally, a bar above each gene/component in the alcoholism pathway indicates its involvement in dopamine, GABA, or opiate addiction.

Addiction-identified genes and their near-neighbor genes from each of the seven addiction hotspots were mapped onto KEGG cellular pathways. We found consistent participation of hotspot genes in pathways involved in neurological signal transmission. Figure 4 illustrates the roles that hotspot addiction genes play in canonical addiction pathways. The acute and chronic alcoholism pathways are heavily influenced at the post-

synaptic neuronal cells by genes contained in all seven of the hotspot regions. The pathway contains genes that carry out the two major mechanisms of addiction regulation: neurotransmission remodeling and epigenetic modification. The major pathways involved in dopaminergic and morphine addiction also show participation of genes with neurotransmitter molecular functions and synaptic transmission biological processes. Figure 4 illustrates that genes belonging to different hotspots coordinate to function in the cellular pathways involved in addiction.

Addiction Hotspots Exhibit Allele Frequency Differences in Human Populations

The presence of addiction hotspots on human chromosomes allowed us to identify genomic variation at shared polymorphisms within a hotspot for the HapMap populations. We found significant differences in allelic distributions both in population comparisons and in population-specific versus neutral comparisons. Hotspot polymorphism distributions were then tested against the proposed neutral (with respect to selective force) loci with the nonparametric Kolmogorov-Smirnov test. To visualize these patterns, we generated the heat maps shown in Figure 5 by converting to the $-\log$ of the P values obtained in pairwise HapMap population comparisons at each hotspot. One of the chromosome 6 hotspots (25.7–26.4 Mb) exhibits an East Asian similarity in allele frequencies in polymorphisms in comparison with non-Asian counterparts.

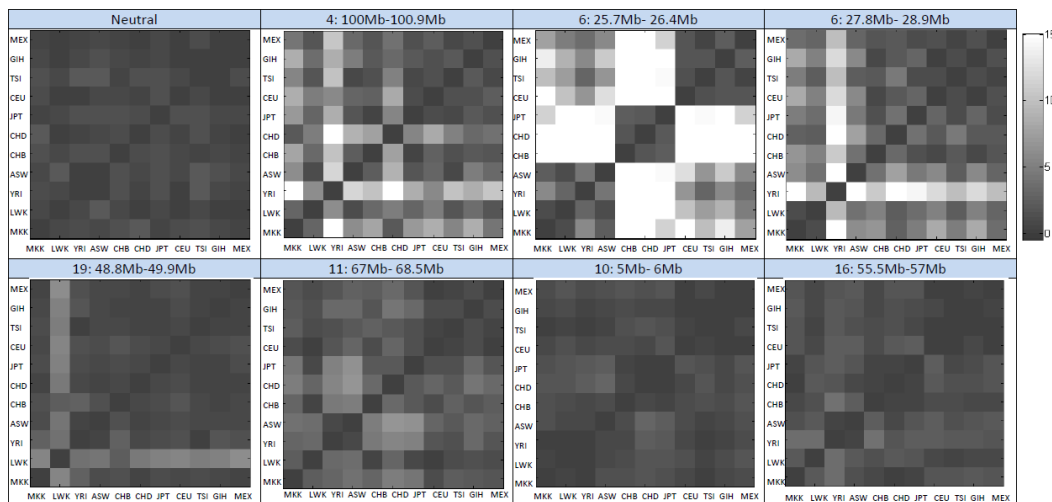


Figure 5. Heat maps of SNP distribution on hotspots show population and geographical patterning. Comparisons of the average reference allele frequencies are shown for each addiction hotspot across populations by using $-\log_{10}$ of the P values obtained from the Kolmogorov-Smirnov test. The variation in neutral regions shows very

little demographic patterning, whereas the hotspot located at Chr6: 25.7-26.4 shows considerably more regional blocks.

Shared polymorphisms at each addiction hotspot were grouped by ethnic geographical origin as Asian, African, or European, with MEX and GIH excluded so that we could make comparisons between these continental-scale ethnic groups. Figure 6 shows the population comparison for polymorphisms along the hotspots. The figure shows that expression differs between Asian, African, and European populations for quite a few polymorphisms in all the addiction hotspots deciphered in this study; the chi-square test ($P < 0.01$) was used in region-level comparisons. Table 2 identifies significant polymorphisms along the chromosome 6 hotspot (25.7 Mb–26.4 Mb) for the three continental population comparisons. A large portion of these polymorphisms fall within the intron regions of histone genes, known to have roles in addiction. However, three others fall within the gene *SLC17A4*, which codes for a sodium phosphate co-transporter in the intestinal mucosa. The protein plays an important role in the absorption of phosphate from the intestine, and its possible role in addiction is yet to be determined. Table 2 also shows five significant intergenic variants identified between these populations (dbSNP entries: rs6906576, rs6924948, rs7740793, rs9348699, and rs933199).

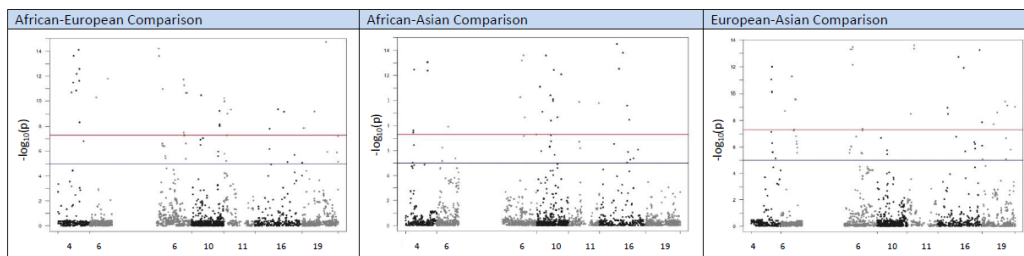


Figure 6. Hotspot SNPs with significantly altered expression in regional population subtypes. Significant polymorphisms were found for regional population comparisons between African-Asian (N=112), African-European (N=126), and European-Asian (N=122) by using the chi-square test with a P value cutoff of $1e-5$, as shown by the blue line in the figure. The value for the red line cutoff is $5e-8$.

Addiction Hotspots show Distant Genomic Interactions

Analyses of addiction hotspots based on allele frequency were complemented with analyses uncovering linkage to adjacent and distal sites in the genome previously identified by GWAS. GWAS3D (M. J. Li et al., 2013) maps interactions between significant polymorphisms previously identified in the “Results” section with annotated genomic interactions culled from hundreds of genotype-phenotype studies and GWAS. Polymorphisms

were characterized as affecting regulatory pathways and underlying disease/trait associations by integrating chromatin state, functional genomics, sequence motif, or conservation information given in genetic polymorphism inputs. Figure 7 shows a projection of the significant common variants between Africans and Europeans onto Yorubans, a sub-Saharan African population from Nigeria. In the outer ring, polymorphisms or genomic region inputs are identified. The second ring localizes the input polymorphism. The thickness of the red lines connecting two polymorphisms indicates the strength of local or long-range interactions.

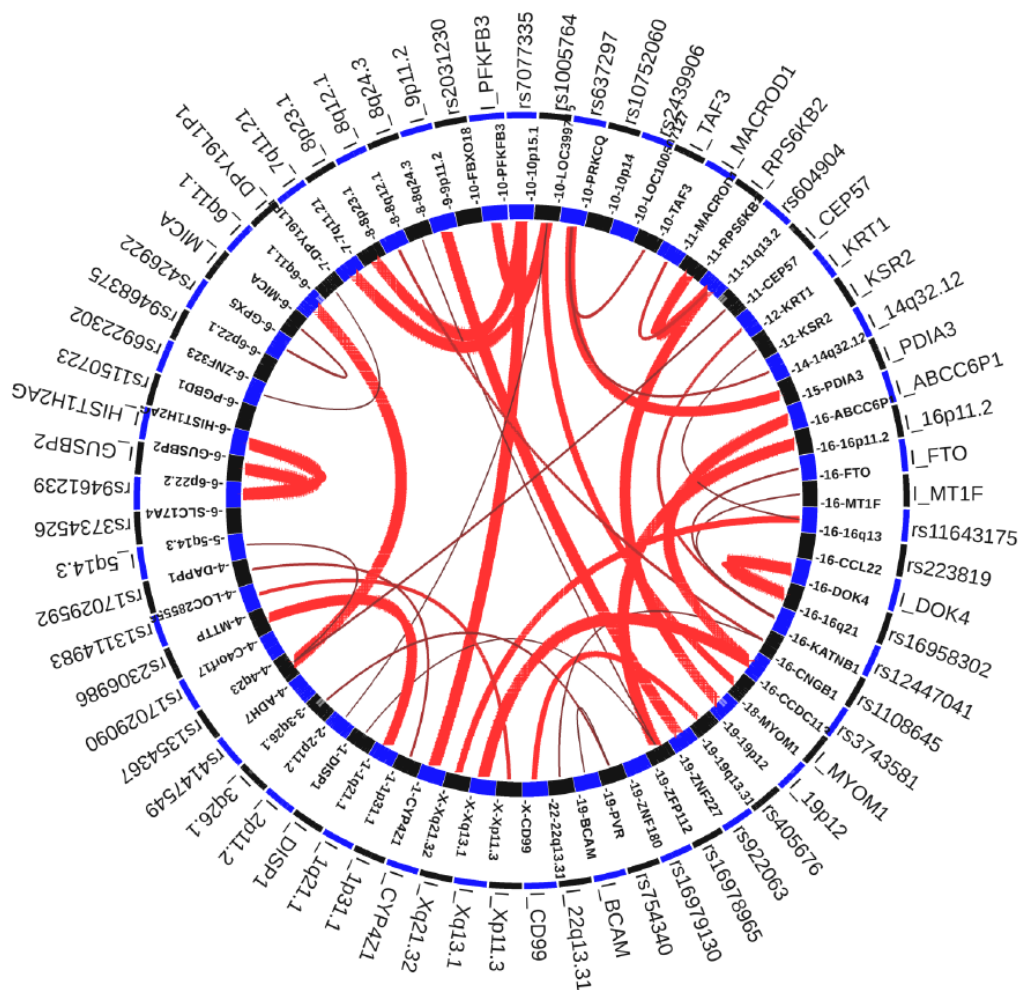


Figure 7. Significant SNPs characterized in the Yoruban population show local and long-range interactions. Polymorphisms identified as significant in African-European comparisons were projected onto the GWAS3D platform. At least six long-range trans-chromosomal interactions were identified, and numerous local interactions were also observed.

SNP	GENE	TYPE	Africans				Asians				European			Other	P values		
			MKK	LWK	YRI	ASW	CHB	CHD	JPT	CEU	TSI	GIH	MEX		African-Asian	African-European	Asian-European
rs1800708	HFE	intron	0.029	0.083	0.088	0.07	0.456	0.445	0.319	0.08	0.05	0.2	0.149	4.04E-01	9.69E-01	1.47E-07	
rs9366637	HFE	Intron	0.045	0.077	0.075	0.079	0.493	0.495	0.385	0.066	0.059	0.173	0.216	3.96E-01	9.17E-01	2.57E-10	
rs2237231	HIST1H1A	downstr. 500B	0.054	0.077	0.075	0.088	0.489	0.5	0.389	0.084	0.059	0.193	0.216	4.01E-01	9.78E-01	5.82E-08	
rs9393682	HIST1H1C	upstr. 2KB	0.042	0.077	0.075	0.079	0.485	0.5	0.389	0.066	0.059	0.203	0.216	3.95E-01	9.27E-01	2.48E-10	
rs2051542	HIST1H1T	missense	0.029	0.082	0.088	0.07	0.46	0.445	0.319	0.08	0.054	0.198	0.164	4.04E-01	9.97E-01	3.59E-07	
rs3830054	HIST1H2AB	upstr. 2KB	0.042	0.077	0.075	0.088	0.496	0.5	0.389	0.084	0.059	0.203	0.216	3.97E-01	9.89E-01	4.85E-08	
rs6908263	HIST1H2AC	intron	0.141	0.114	0.136	0.105	0.46	0.45	0.319	0.08	0.064	0.198	0.172	4.86E-01	4.70E-01	2.73E-06	
rs7760713	HIST1H2AC	intron	0.138	0.114	0.139	0.105	0.463	0.449	0.321	0.076	0.064	0.193	0.216	4.85E-01	4.40E-01	1.11E-06	
rs9467684	HIST1H2BD	intron	0.179	0.132	0.153	0.149	0.011	0.005	0.08	0.013	0.025	0.035	0.017	1.51E-04	1.60E-12	4.94E-01	
rs4145878	HIST1H2BF	upstr.2K B intron	0.391	0.405	0.412	0.404	0.106	0.092	0.018	0.5	0.495	0.381	0.414	4.28E-06	8.49E-01	3.92E-01	
rs1892252	SLC17A4	intron	0.362	0.232	0.248	0.316	0.051	0.032	0.075	0.133	0.098	0.03	0.172	6.90E-06	1.32E-01	5.86E-01	
rs3734525	SLC17A4	utr 3'	0.106	0.186	0.211	0.193	0.007	0.005	0.075	0.004	0.01	0.005	0.017	5.73E-07	0.00E+00	1.67E-03	
rs3823151	SLC17A4	intron	0.388	0.336	0.35	0.272	0.099	0.087	0.111	0.04	0.049	0.03	0.078	1.64E-02	5.32E-11	2.21E-01	
rs199738	TRIM38	utr 5'	0.337	0.403	0.445	0.36	0.074	0.046	0.053	0.19	0.157	0.183	0.147	1.21E-08	2.20E-01	5.04E-01	
rs6906576	intergenic	genomic	0.074	0.114	0.175	0.14	0.391	0.408	0.35	0.058	0.039	0.183	0.241	5.02E-01	1.11E-01	5.31E-12	
rs6924948	intergenic	genomic	0.023	0.028	0.028	0.035	0.485	0.495	0.386	0.076	0.054	0.215	0.207	3.49E-01	5.74E-01	1.91E-09	
rs7740793	intergenic	genomic	0.122	0.114	0.19	0.149	0.391	0.408	0.35	0.058	0.039	0.175	0.241	5.32E-01	4.95E-02	5.31E-12	
rs9348699	intergenic	genomic	0.122	0.114	0.19	0.149	0.391	0.408	0.35	0.058	0.039	0.183	0.241	5.32E-01	4.95E-02	5.31E-12	
rs933199	intergenic	genomic	0.03	0.083	0.091	0.07	0.459	0.445	0.319	0.082	0.054	0.203	0.17	4.05E-01	9.94E-01	5.88E-07	

Table 2. SNPs identified as significant in comparisons between Africans, Asians, and Europeans for the 6p21.2 addiction hotspot. All SNPs were tested for significant differences in allele frequencies by using a chi-square test (cutoff <1e-5) in pairwise regional comparisons: Africans (ASW, YRI, LWK, and MKK); Europeans (CEU and TSI); and Asians (CHB, CHD, JPT). The allele frequencies of variants in populations are given with the *P* values of the grouped regional ethnic populations.

Discussion

These results identify for the first time the presence of seven opiate, dopamine, and GABA hotspots across the genome. All but two of these hotspots share functional annotation between those genes previously identified from the addiction literature as participating in addiction phenotypes and their co-located gene neighbors. Furthermore, when all genes within a hotspot window are mapped onto KEGG pathways, both canonical pathways, such as acute alcohol intoxication, and more unconventional pathways, such as those involved in systemic lupus, are identified. Because addiction phenotypes show ethnic population specificity, publicly available polymorphisms from 11 populations were assessed at each hotspot location. Three striking ethnicity-based signatures arose, at the chromosome 4 hotspot and at one of the chromosome 6 hotspots; the Yorubans, a sub-Saharan population of Africans, showed differences from all other ethnic populations. Meanwhile, at the other chromosome 6 hotspot, East Asian populations showed differences from all other surveyed populations. These findings suggest that East Asian populations may have both genetic and epigenetic variation for alcohol metabolic processes. Genomic polymorphisms were compared with those in other GWAS identified studies and found to differ from their genomic neighbors in allelic frequencies. Histone gene families are also critical to epigenomic processes, representing an additional layer of substrate for ethnic differences in human populations. Finally, comparisons of significant polymorphisms in this analysis with those identified in other GWAS generated a slew of long-range trans-chromosomal linkages to well-characterized addiction genes, such as the monoamine oxidase A gene (*MAOA*). The enzyme monoamine oxidase catalyzes metabolic reactions involving dopamine, norepinephrine, and serotonin.

Other studies have identified genomic regions of gene set enrichment for particularly complex phenotypes (Bradley et al., 2010). About 11% of the literature-curated addiction genes formed seven clusters of six or more genes within a span of 1.5 Mb, the average distance of recombination in humans. This linear genomic distance was chosen to attempt to find those regions that are highly likely to be inherited from one generation to the next. Three of the seven addiction hotspots (on chromosomes 4 and 6) were related exclusively to GABA addiction. The rest contained genes that participated in addiction to all three classes of illicit substances. Because the hotspot genes also participated in numerous metabolic pathways unrelated to addiction, such as the pathways for systemic lupus erythematosus, viral carcinogenesis, and steroid hormone

biosynthesis, we suggest that they may confer susceptibilities to additional disorders in long-term substance abusers. The functionally pleiotropic nature of these hotspots may help to explain how substance addiction creates a spectrum of disorders that are not limited to a particular addictive substance (Beitner-Johnson & Nestler, 1991; Schmidt, McGinty, West, & Sadri-Vakili, 2013).

The presence of hotspots on the human genome provides an opportunity for the discovery of addiction-related genes. Addiction genes accounted for a significant percentage of the total number of genes (>36%) in the three hotspots exclusively related to GABA addiction. The four mixed hotspots contained many genes not currently linked to addiction. Consider, for example, the gene *RG9MTD2* in the GABA-only hotspot on chromosome 4, which codes for an RNA transmethylase expressing an acetaminophen-binding site. Acetaminophen has been shown to increase feelings of intoxication in combination with ethanol in a human cohort and does not mitigate subjective feelings of alcohol intoxication (Pickworth, Klein, George, & Henningfield, 1992). Additionally, our mixed addiction hotspot located at chromosome 19 is adjacent to the killer cell immunoglobulin-like receptor (*KIR*) genomic region. This locus displays extensive diversity through polymorphism within individual *KIR* genes (Hsu, Chida, Geraghty, & Dupont, 2002). We speculate that the polymorphisms at this locus could promote regulatory changes for the *KIR* region. Another family of genes with potentially important roles, still undefined, are the *SLC17A1-4* genes, which code for organic ion transporters in a GABA-only hotspot on chromosome 6. These genes were identified as having storage activity (Reimer & Edwards, 2004) for neurotransmitters and therefore might be playing crucial roles in addiction (Tomkins & Sellers, 2001).

Addiction hotspots show polymorphism differences in ethnic populations. Pooling these populations according to their geographical origins (East Asia, Africa, and Europe) allowed a continental perspective of the importance of ethnic origin in future analyses of genetic and epigenetic interactions. Our analyses found significant population variation at a chromosome 6 location, which differentiated HapMap East Asian populations from all non-Asian populations. This type of assessment provides a potentially potent source of genetic comparisons that could help explain phenotype differences between East Asian and non-Asian populations. Additionally, the Yoruban (Nigerian) population exhibits unique signals of variation at chromosomes 4 and 6. Additional analyses of GWAS identifying addiction polymorphisms in the 11 HapMap populations provide support for the allele frequency patterns reported in our analysis.

The gene contents of GABA-specific addiction hotspots suggest an epigenetic role in alcohol addiction (Bali, Im, & Kenny, 2011; Glahn et al., 2013; Miller, Campbell, & Sweatt, 2008; Ponomarev, 2013). The two hotspots on chromosome 6 are highly enriched with histone genes. It is well established that the methylation state of histone proteins is directly related to whether DNA tracts are turned off or on. Functional annotation shows histone involvement in KEGG pathways unrelated to addiction metabolism, such as those for systemic lupus, viral carcinogenesis, and transcriptional misregulation of cancer. Histone protein modifications through methylation have been identified as one of the possible reasons for the diverse phenotypes observed in substance-addicted individuals (Bilinski et al., 2012; Schifano, Li, Christiani, & Lin, 2013; Schmidt et al., 2013). The polymorphisms in the histone genes on chromosome 6 may result in alternate epigenetic modifications among ethnic populations, a result previously seen in other data sets (Fraser, Lam, Neumann, & Kobor, 2012). Moreover, our GWAS3D analyses found that the chromosome 6 polymorphisms identified as significant in our cross-population assessments were also involved in local and trans-chromosomal interactions. Potential links between hotspots, addiction genes, and differences in polymorphism frequency merit further investigation. This analysis leverages a systems biology framework to incorporate disparate sources of data. Currently, data such as those garnered from the human HapMap project, ClinVar, GWAS Central, and dbSNP provide rich source materials for the creation of genetic predictive models to identify at-risk communities and ethnic populations within the United States. A clear challenge continues to be the integration of sociological and environmental data with existing genomic data analyses. Nevertheless, this approach provides a foundation for expanding interdisciplinary data analysis, so that greater inferences can be made in understanding complex phenotypes, such as those of opiate, dopamine, and alcohol substance addiction; it also has the potential to accelerate the development of precision medicine models for vulnerable and at-risk groups.

REFERENCES

- Adkins, D. E., McClay, J. L., Vunck, S. A., Batman, A. M., Vann, R. E., Clark, S. L., ... Beardsley, P. M. (2013). Behavioral metabolomics analysis identifies novel neurochemical signatures in methamphetamine sensitization. *Genes Brain and Behavior*, 12(8), 780–791. doi:10.1111/gbb.12081
- Bali, P., Im, H. I., & Kenny, P. J. (2011). Methylation, memory and addiction. *Epigenetics*, 6(6), 671–674.
- Beitner-Johnson, D., & Nestler, E. J. (1991). Morphine and cocaine exert common chronic actions on tyrosine hydroxylase in dopaminergic brain reward regions. *Journal of Neurochemistry*, 57(1), 344–347.
- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research*, 125(1–2), 279–284.
- Bilinski, P., Wojtyla, A., Kapka-Skrzypczak, L., Chwedorowicz, R., Cyranka, M., & Studzinski, T. (2012). Epigenetic regulation in drug addiction. *Annals of Agricultural and Environmental Medicine*, 19(3), 491–496.
- Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., ... Taylor, J. (2010). Galaxy: A web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology*, Chapter 19, Unit 19.10.1–21. doi:10.1002/0471142727.mb1910s89
- Bradley, W. E., Raelson, J. V., Dubois, D. Y., Godin, E., Fournier, H., Prive, C., ... Belouchi, A. (2010). Hotspots of large rare deletions in the human genome. *PLoS One*, 5(2), e9401. doi:10.1371/journal.pone.0009401
- Brown, J. N., Ortiz, G. M., Angel, T. E., Jacobs, J. M., Gritsenko, M., Chan, E. Y., ... Smith, R. D. (2012). Morphine produces immunosuppressive effects in nonhuman primates at the proteomic and cellular levels. *Molecular and Cellular Proteomics*, 11(9), 605–618. doi:10.1074/mcp.M111.016121
- Chao, J., & Nestler, E. J. (2004). Molecular neurobiology of drug addiction. *Annual Review of Medicine*, 55, 113–132. doi:10.1146/annurev.med.55.091902.103730
- Chiang, Y. C., Lo, Y. N., & Chen, J. C. (2013). Crosstalk between dopamine D(2) receptors and cannabinoid CB(1) receptors regulates CNR1 promoter activity via ERK1/2 signaling. *Journal of Neurochemistry*, 127(2), 163–176. doi:10.1111/jnc.12399

- Commons, K. G. (2010). Neuronal pathways linking substance P to drug addiction and stress. *Brain Research*, 1314, 175–182. doi:10.1016/j.brainres.2009.11.014
- Contet, C. (2012). Gene expression under the influence: Transcriptional profiling of ethanol in the brain. *Current Psychopharmacology*, 1(4), 301–314.
- Cotton, R. W., & Goldman, D. (1988). Review of the molecular biology of the human alcohol dehydrogenase genes and gene products. *Advances in Alcohol & Substance Abuse*, 7(3-4), 171–182. doi:10.1300/J251v07n03_22
- Crabb, D. W., Bosron, W. F., & Li, T. K. (1987). Ethanol metabolism. *Pharmacology & Therapeutics*, 34(1), 59–73.
- Cunningham, J. A., Sobell, L. C., & Chow, V. M. (1993). What's in a label? The effects of substance types and labels on treatment considerations and stigma. *Journal of Studies on Alcohol*, 54(6), 693–699.
- Dean, J. C., & Rud, F. (1984). The drug addict and the stigma of addiction. *The International Journal of the Addictions*, 19(8), 859–869.
- Ehlers, C. L., Liang, T., & Gizer, I. R. (2012). ADH and ALDH polymorphisms and alcohol dependence in Mexican and Native Americans. *The American Journal of Drug and Alcohol Abuse*, 38(5), 389–394. doi:10.3109/00952990.2012.694526
- Fraser, H. B., Lam, L. L., Neumann, S. M., & Kobor, M. S. (2012). Population-specificity of human DNA methylation. *Genome Biology*, 13(2), R8. doi:10.1186/gb-2012-13-2-r8
- Freeman, W. M., Brebner, K., Amara, S. G., Reed, M. S., Pohl, J., & Phillips, A. G. (2005). Distinct proteomic profiles of amphetamine self-administration transitional states. *Pharmacogenomics Journal*, 5(3), 203–214. doi:10.1038/sj.tpj.6500309
- Fuerst, P. A., Chakraborty, R., & Nei, M. (1977). Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. *Genetics*, 86(2 Pt. 1), 455–483.
- Gardner, E. L. (2011). Addiction and brain reward and antireward pathways. *Advances in Psychosomatic Medicine*, 30, 22–60. doi:10.1159/000324065
- Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., ... Nekrutenko, A. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, 15(10), 1451–1455. doi:10.1101/gr.4086505

- Girard, D. E., & Carlton, B. E. (1978). Alcoholism. Earlier diagnosis and definition of the problem. *The Western Journal of Medicine*, 129(1), 1–7.
- Glahn, A., Riera Knorrenschild, R., Rhein, M., Haschemi Nassab, M., Groschl, M., Heberlein, A., ... Hillemacher, T. (2013). Alcohol-induced changes in methylation status of individual CpG sites, and serum levels of vasopressin and atrial natriuretic peptide in alcohol-dependent patients during detoxification treatment. *European Addiction Research*, 20(3), 143–150. doi:10.1159/000357473
- Goecks, J., Nekrutenko, A., Taylor, J., & Galaxy, T. (2010). Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11(8), R86. doi:10.1186/gb-2010-11-8-r86
- Goodman, A. (1990). Addiction: Definition and implications. *British Journal of Addiction*, 85(11), 1403–1408.
- Goodwin, D. W. (1975). Genetic determinants of alcohol addiction. *Advances in Experimental Medicine and Biology*, 56, 339–355.
- Hillemacher, T. (2011). Biological mechanisms in alcohol dependence—new perspectives. *Alcohol and Alcoholism*, 46(3), 224–230. doi:10.1093/alcalc/agr026
- Hoffman, R. S., & Goldfrank, L. R. (1990). The impact of drug abuse and addiction on society. *Emergency Medicine Clinics of North America*, 8(3), 467–480.
- Hsu, K. C., Chida, S., Geraghty, D. E., & Dupont, B. (2002). The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. *Immunological Reviews*, 190, 40–52.
- Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57. doi:10.1038/nprot.2008.211
- Huang da, W., Sherman, B. T., Zheng, X., Yang, J., Imamichi, T., Stephens, R., & Lempicki, R. A. (2009). Extracting biological meaning from large gene lists with DAVID. *Current Protocols in Bioinformatics*, Chapter 13, Unit 13.11. doi:10.1002/0471250953.bi1311s27
- International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, 437(7063), 1299–1320. doi:10.1038/nature04226
- International HapMap Consortium, Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., ... McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse

- human populations. *Nature*, 467(7311), 52–58.
doi:10.1038/nature09298
- International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., ... Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164), 851–861. doi:10.1038/nature06258
- Kanehisa, M. (2002). The KEGG database. *Novartis Foundation Symposium*, 247, 91–101; discussion 101–103, 119–128, 244–152.
- Kanehisa, M., Goto, S., Kawashima, S., & Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Research*, 30(1), 42–46.
- Kreek, M. J., Nielsen, D. A., & LaForge, K. S. (2004). Genes associated with addiction: alcoholism, opiate, and cocaine addiction. *Neuromolecular Medicine*, 5(1), 85–108. doi:10.1385/NMM:5:1:085
- Le Foll, B., Gallo, A., Le Strat, Y., Lu, L., & Gorwood, P. (2009). Genetics of dopamine receptors and drug addiction: a comprehensive review. *Behavioural Pharmacology*, 20(1), 1–17.
doi:10.1097/FBP.0b013e3283242f05
- Lee, A. M., & Messing, R. O. (2008). Protein kinases and addiction. *Annals of the New York Academy of Sciences*, 1141, 22–57.
doi:10.1196/annals.1441.022
- Li, C. Y., Mao, X., & Wei, L. (2008). Genes and (common) pathways underlying drug addiction. *PLoS Computational Biology*, 4(1), e2.
doi:10.1371/journal.pcbi.0040002
- Li, K. W., Jimenez, C. R., van der Schors, R. C., Hornshaw, M. P., Schoffemeer, A. N., & Smit, A. B. (2006). Intermittent administration of morphine alters protein expression in rat nucleus accumbens. *Proteomics*, 6(6), 2003–2008.
doi:10.1002/pmic.200500045
- Li, M. J., Wang, L. Y., Xia, Z., Sham, P. C., & Wang, J. (2013). GWAS3D: Detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. *Nucleic Acids Research*, 41(Web Server issue), W150–W158. doi:10.1093/nar/gkt456
- Lorman, W. J. (2013). Maintaining sobriety and recovery. *The Nursing Clinics of North America*, 48(3), 437–444, vi.
doi:10.1016/j.cnur.2013.04.005
- Lynskey, M. T., & Strang, J. (2013). The global burden of drug use and mental disorders. *Lancet*, 382(9904), 1540–1542.
doi:10.1016/S0140-6736(13)61781-X
- Maldonado, R. (2003). The neurobiology of addiction. *Journal of Neural Transmission. Supplementum*(66), 1–14.

- Messer, C. J., Eisch, A. J., Carlezon, W. A. Jr., Whisler, K., Shen, L., Wolf, D. H., ... Nestler, E. J. (2000). Role for GDNF in biochemical and behavioral adaptations to drugs of abuse. *Neuron*, 26(1), 247–257.
- Miller, C. A., Campbell, S. L., & Sweatt, J. D. (2008). DNA methylation and histone acetylation work in concert to regulate memory formation and synaptic plasticity. *Neurobiology of Learning and Memory*, 89(4), 599–603. doi:10.1016/j.nlm.2007.07.016
- Mulligan, M. K., Ponomarev, I., Hitzemann, R. J., Belknap, J. K., Tabakoff, B., Harris, R. A., ... Bergeson, S. E. (2006). Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 103(16), 6368–6373. doi:10.1073/pnas.0510188103
- Naranjo, C. A., & Bremner, K. E. (1993). Behavioural correlates of alcohol intoxication. *Addiction*, 88(1), 25–35.
- Olfson, E., & Bierut, L. J. (2012). Convergence of genome-wide association and candidate gene studies for alcoholism. *Alcoholism, Clinical and Experimental Research*, 36(12), 2086–2094. doi:10.1111/j.1530-0277.2012.01843.x
- Pasche, B., & Yi, N. (2010). Candidate gene association studies: Successes and failures. *Current Opinion in Genetics & Development*, 20(3), 257–261. doi:10.1016/j.gde.2010.03.006
- Peele, S. (1977). Redefining addiction. I. Making addiction a scientifically and socially useful concept. *International Journal of Health Services*, 7(1), 103–124.
- Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature*, 465(7299), 721–727. doi:10.1038/nature09230
- Pickworth, W. B., Klein, S. A., George, F. R., & Henningfield, J. E. (1992). Acetaminophen fails to inhibit ethanol-induced subjective effects in human volunteers. *Pharmacology, Biochemistry, and Behavior*, 41(1), 189–194.
- Pietruszko, R. (1975). Mammalian liver alcohol dehydrogenases. *Advances in Experimental Medicine and Biology*, 56, 1–31.
- Ponomarev, I. (2013). Epigenetic control of gene expression in the alcoholic brain. *Alcohol Research*, 35(1), 69–76.
- Popendikyte, V., Laurinavicius, A., Paterson, A. D., Macciardi, F., Kennedy, J. L., & Petronis, A. (1999). DNA methylation at the putative promoter region of the human dopamine D2 receptor gene. *Neuroreport*, 10(6), 1249–1255.

- Reimer, R. J., & Edwards, R. H. (2004). Organic anion transport is the primary function of the SLC17/type I phosphate transporter family. *Pflugers Archiv*, 447(5), 629–635. doi:10.1007/s00424-003-1087-y
- Room, R. (2005). Stigma, social inequality and alcohol and drug use. *Drug and Alcohol Review*, 24(2), 143–155. doi:10.1080/09595230500102434
- Saito, M., Szakall, I., Toth, R., Kovacs, K. M., Oros, M., Prasad, V. V., ... Vadasz, C. (2004). Mouse striatal transcriptome analysis: effects of oral self-administration of alcohol. *Alcohol*, 32(3), 223–241. doi:10.1016/j.alcohol.2004.02.005
- Schifano, E. D., Li, L., Christiani, D. C., & Lin, X. (2013). Genome-wide association analysis for multiple continuous secondary phenotypes. *American Journal of Human Genetics*, 92(5), 744–759. doi:10.1016/j.ajhg.2013.04.004
- Schmidt, H. D., McGinty, J. F., West, A. E., & Sadri-Vakili, G. (2013). Epigenetics and psychostimulant addiction. *Cold Spring Harbor Perspectives in Medicine*, 3(3), a012047. doi:10.1101/cshperspect.a012047
- Schumacher, A., & Petronis, A. (2006). Epigenetics of complex diseases: From general theory to laboratory experiments. *Current Topics in Microbiology and Immunology*, 310, 81–115.
- Shi, J., Hui, L., Xu, Y., Wang, F., Huang, W., & Hu, G. (2002). Sequence variations in the mu-opioid receptor gene (OPRM1) associated with human addiction to heroin. *Human Mutation*, 19(4), 459–460. doi:10.1002/humu.9026
- Sun, J., & Zhao, Z. (2010). Functional features, biological pathways, and protein interaction networks of addiction-related genes. *Chemistry & Biodiversity*, 7(5), 1153–1162. doi:10.1002/cbdv.200900319
- Tomkins, D. M., & Sellers, E. M. (2001). Addiction and the brain: The role of neurotransmitters in the cause and treatment of drug dependence. *Canadian Medical Association Journal*, 164(6), 817–821.
- Treutlein, J., & Rietschel, M. (2011). Genome-wide association studies of alcohol dependence and substance use disorders. *Current Psychiatry Reports*, 13(2), 147–155. doi:10.1007/s11920-011-0176-4
- Truan, F. (1993). Addiction as a social construction: A postempirical view. *The Journal of Psychology*, 127(5), 489–499. doi:10.1080/00223980.1993.9914886
- Uhl, G. R., Drgon, T., Johnson, C., Li, C. Y., Contoreggi, C., Hess, J., ... Liu, Q. R. (2008). Molecular genetics of addiction and related

- heritable phenotypes: Genome-wide association approaches identify "connectivity constellation" and drug target genes with pleiotropic effects. *Annals of the New York Academy of Sciences*, 1141, 318–381. doi:10.1196/annals.1441.018
- Wall, J. D., Cox, M. P., Mendez, F. L., Woerner, A., Severson, T., & Hammer, M. F. (2008). A novel DNA sequence database for analyzing human demographic history. *Genome Research*, 18(8), 1354–1361. doi:10.1101/gr.075630.107
- Yuferov, V., Levran, O., Proudnikov, D., Nielsen, D. A., & Kreek, M. J. (2010). Search for genetic markers and functional variants involved in the development of opiate and cocaine addiction and treatment. *Annals of the New York Academy of Sciences*, 1187, 184–207. doi:10.1111/j.1749-6632.2009.05275.x
- Zhang, G. C., Vu, K., Parelkar, N. K., Mao, L. M., Stanford, I. M., Fibuch, E. E., & Wang, J. Q. (2009). Acute administration of cocaine reduces metabotropic glutamate receptor 8 protein expression in the rat striatum in vivo. *Neuroscience Letters*, 449(3), 224–227. doi:10.1016/j.neulet.2008.11.008
- Zhang, R., Miao, Q., Wang, C., Zhao, R., Li, W., Haile, C. N., ... Zhang, X. Y. (2013). Genome-wide DNA methylation analysis in alcohol dependence. *Addiction Biology*, 18(2), 392–403. doi:10.1111/adb.12037